UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/695,596	10/28/2003	Phillip Jeffrey Bloom	SYAR-100	5364

7590 08/18/2008

STALLMAN & POLLOCK LLP
Attn: Michael A. Stallman
Suite 2200
353 Sacramento Street
San Francisco, CA 94111

EXAMINER	
FLETCHER, JAMES A	
ART UNIT	PAPER NUMBER
2621	

MAIL DATE	DELIVERY MODE
08/18/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/695,596	BLOOM ET AL.	
	Examiner	Art Unit	
	JAMES A. FLETCHER	2621	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *04 August 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-30* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-30* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _____ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *02/04 08/04*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

New Art Unit

1. Please include the new Art Unit 2621 in the caption or heading of any written or facsimile communication submitted after this Office Action because the examiner, who was assigned to Art Unit 2616, has been assigned to new Art Unit 2621. Your cooperation in this matter will assist in the timely processing of the submission and is appreciated by the Office.

Claim Rejections - 35 USC § 101

2. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 16-18, 22-24, and 30 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. In particular, the cited claims recite "software" without the enabling limitation of being provided on a computer-readable medium. Appropriate correction is required.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 1-5, and 10-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman et al (6,260,011) and further in view of Nefian (7,165,029).

Regarding claim 1, Heckerman et al disclose an audio and text data processor comprising:

- a selector for selecting at least a portion of an audio data stream (Col. 5, lines 41-42 "A plurality of M audio files 22, 24 form an audio corpus 20");
- an audio feature analyser for abstracting from said selected portion of said audio data stream a stream of time-varying features and for abstracting corresponding time-varying features from an input audio data stream (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");
- a timing analysis and waveform editing processor adapted to determine timing differences between said stream of time-varying features and said corresponding time-varying features and to utilize said timing differences to edit said input audio data stream (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus"); and

- a playback control module adapted to control running of said synchronized audio data and video data streams with said edited input audio data stream replacing said selected portion (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Regarding claim 2, Heckerman et al disclose a data processing system for audio and video data, comprising:

- digitized audio and video data for providing an audio data stream synchronized with a data stream (Col. 5, lines 41-42 “A plurality of M audio files 22, 24 form an audio corpus 20”);
- timing data representative of a plurality of selected times in a running of said synchronized audio and video data streams (Col 8, lines 28-33 “The alignment module 318 is also responsible for aligning the audio and text files based on the identified alignment points, e.g., by inserting into the text and/or audio files time stamps or other markers which can be used as pointers between the audio and text files”);
- audio feature data for providing a data stream of time-varying features abstracted from at least a selected portion of said audio data stream (Col 11, lines 26-30 “locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize”);
- an audio feature analyser for abstracting a corresponding stream of time-varying features from an input audio data stream (Col 11, lines 26-30 “locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and

recorded for commercial distribution purposes, silence is often a particularly easy to recognize");

- a timing analysis and waveform editing processor adapted to determine timing differences between said streams of time-varying features and to utilize said timing differences to edit said input audio data stream and produce edited input audio data (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus"); and
- a playback control module adapted to control running said synchronized audio data and data streams with said edited input audio data replacing said selected portion (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files").

Further regarding claim 2, please see Examiner's remarks regarding claim 1 above.

Regarding claim 3, Heckerman et al disclose a data processing system comprising cueing data representative of timing of said selected portion of said audio data stream (Col 8, lines 28-33 "The alignment module 318 is also responsible for aligning the audio and text files based on the identified alignment points, e.g., by inserting into the text and/or audio files time stamps or other markers which can be used as pointers between the audio and text files").

Regarding claim 4, Heckerman et al disclose a data processing system comprising additional digitized audio data for providing a further audio data stream synchronized with said data stream (Col 8, lines 16-18 "The language model generation module 314 is used for generating, from a text corpus, a language model used by the speech recognizer 312").

Regarding claim 5, Heckerman et al disclose a process for providing audio and video data, comprising the steps of:

- selecting at least a portion of said audio data stream (Col. 5, lines 41-42 "A plurality of M audio files 22, 24 form an audio corpus 20");
- analysing said selected portion to abstract therefrom a stream of time-varying features (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize"); and
- providing control data relating said selected portion to said stream of time-varying features (Col 8, lines 28-33 "The alignment module 318 is also responsible for aligning the audio and text files based on the identified alignment points, e.g., by inserting into the text and/or audio files time stamps or other markers which can be used as pointers between the audio and text files").

Further regarding claim 5, please see Examiner's remarks regarding claim 1 above.

Regarding claim 11, Heckerman et al disclose a process wherein more than one portion of said audio data stream is selected (Col 13, lines 40-42 "pointers, synchronizing portions of the audio and text data").

Regarding claim 10, Heckerman et al disclose a method of processing audio data, comprising the steps of:

- selecting at least a portion of said original audio data stream (Col. 5, lines 41-42 "A plurality of M audio files 22, 24 form an audio corpus 20");
- storing an input audio data stream substantially in synchronization with a portion of said data stream corresponding to the selected portion of said original audio data stream (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word");
- abstracting from said input audio data stream a stream of time-varying features of the input audio data stream (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");

- comparing the abstracted stream of time-varying features with a corresponding stream of time-varying features abstracted from said selected portion of said original audio data stream and determining timing differences between said streams of time-varying features (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus");
- utilizing said timing differences to edit said input audio data stream and produce edited input audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files"); and
- running said synchronized original audio data stream and video data stream with said edited input audio data replacing said selected portion (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files").

Further regarding claim 10, please see Examiner's remarks regarding claim 1 above.

Regarding claim 12, Heckerman et al disclose a method according to claim 10, wherein more than one portion of said original audio data stream is selected (Col 13, lines 40-42 "pointers, synchronizing portions of the audio and text data").

Regarding claim 13, Heckerman et al disclose an apparatus for processing audio data, comprising:

- means for deriving from audio data feature data representative of audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");
- means for comparing first feature data derived from first audio data synchronously associated with data with second feature data derived from second audio data and determining timing differences between the first and second feature data (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus");
- means for editing the second audio data in dependence upon said timing difference such as to provide edited second audio data in a synchronous relation to said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files"); and
- means for synchronously outputting said video data and said edited second audio data while muting said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer

linking the text and audio files, to the text files, and, optionally, to the audio files").

Further regarding claim 13, please see Examiner's remarks regarding claim 1 above.

Regarding claim 14, Heckerman et al disclose an apparatus for processing audio data, comprising:

- means for deriving from audio data feature data representative of audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");
- means for selecting from data representing synchronously streamable video and audio data, data representing a portion of a stream of the streamable data and measuring durations of and intervals containing audible time-varying acoustic features of the audio data (Col 11, lines 39-2 "for a pointer to be inserted into the text and/or audio for synchronization purposes, the recognized text, bracketing the identified point of silence must have been correctly identified"); and
- means for populating a database with data and measurements provided by said selecting and measuring means (Col 9, lines 1-5 "The speech recognizer

module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Further regarding claim 14, please see Examiner's remarks regarding claim 1 above.

Regarding claim 15, Heckerman et al disclose an apparatus comprising means for populating said database with text related to said data and measurements provided by said selecting and measuring means (Col 8, lines 14-16 "the speech recognizer module 312 generates a set of recognized text with time stamps from one or more audio files").

Regarding claim 16, Heckerman et al disclose an audio and video data processing software (Col 5, line 66 – Col 6, line 1 "The present invention will be described in the general context of computer-executable instructions, such as program modules, being executed by a personal computer") comprising:

- a feature analysis program adapted to derive from audio data feature data representative of audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");

- a comparison and timing program adapted to compare first feature data derived from first audio data synchronously associated with data with second feature data derived from second audio data and to determine timing differences between the first and second feature data (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus");
- an editing program adapted to edit the second audio data in dependence upon said timing differences such as to provide edited second audio data in a synchronous relation to said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files"); and
- a streaming program adapted to synchronously, output said video data and said edited second audio data while muting said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files").

Further regarding claim 16, please see Examiner's remarks regarding claim 1 above.

Regarding claim 17, Heckerman et al disclose audio and video data processing software (Col 5, line 66 – Col 6, line 1 "The present invention will be described in the

general context of computer-executable instructions, such as program modules, being executed by a personal computer") comprising:

- a feature analysis program adapted to derive from audio data feature data representative of audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");
- a selection and measuring program adapted to select from data representing synchronously streamable video and audio data, data representing a portion of a stream of the streamable data and to measure durations of and intervals containing audible time-varying acoustic features of the audio data (Col 11, lines 39-2 "for a pointer to be inserted into the text and/or audio for synchronization purposes, the recognized text, bracketing the identified point of silence must have been correctly identified"); and
- a database program adapted to populate a database with data and measurements provided by said selection and measuring program (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Regarding claim 18, Heckerman et al disclose audio and video data processing software according to claim 17, wherein said database program is further adapted to enable population of said database with text related to said data and measurements provided by said selection and measuring program (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

5. Claims 6 and 7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman et al and Nefian, and further in view of Okada et al (5,809,454).

Regarding claim 6, Heckerman et al disclose an method of providing a processing system for audio and video data, comprising the steps of:

- storing timing data representative of a plurality of selected times in a running of said synchronized audio and data streams (Col 8, lines 28-33 "The alignment module 318 is also responsible for aligning the audio and text files based on the identified alignment points, e.g., by inserting into the text and/or audio files time stamps or other markers which can be used as pointers between the audio and text files");
- selecting at least a portion of said audio data stream (Col. 5, lines 41-42 "A plurality of M audio files 22, 24 form an audio corpus 20");
- abstracting from the selected portion of said audio data stream audio feature data for providing a data stream of time-varying features (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by

correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize"); and

- storing a playback control module for controlling running said synchronized audio data and video data streams with edited input audio data from said processor replacing said selected portion (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files");
- storing the abstracted audio feature data (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word");
- storing an audio feature analyser for abstracting a corresponding stream of time-varying features from an input audio data stream (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word");

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a

display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Heckerman discloses a processor adapted to determine timing between two related streams of data (Col 13, lines 40-42 "pointers, synchronizing portions of the audio and text data which have been found to correspond to each other"), but do not specifically disclose an editing processor to determine timing differences between the streams.

Okada et al teach storage of a processor (Col 21, lines 43-46 "the signal processing in individual circuits 1 to 55 may be replaced with software-based signal processing which is accomplished by using a CPU") adapted to determine timing differences between data streams (Col 7, lines 45-47 "The speech length compressor/expander 43 compresses or expands the sound interval determined

by the voice determining circuit 41"), providing the user with sound synchronized to image when the playback rate is modified.

As taught by Okada, storage of a processor adapted to sense timing differences between streams is well known, and would therefore have been obvious to one of ordinary skill in the art at the time of the invention to modify Heckerman accordingly.

Regarding claim 7, Heckerman et al disclose a method comprising the step of: storing cueing data representative of timing of said selected portion of said audio data stream (Col 8, lines 28-33 "The alignment module 318 is also responsible for aligning the audio and text files based on the identified alignment points, e.g., by inserting into the text and/or audio files time stamps or other markers which can be used as pointers between the audio and text files").

6. Claim 8 is rejected under 35 U.S.C. 103(a) as being unpatentable over the combination as applied to claim 6 above, and further in view of Wakamoto (6,283,760).

Regarding claim 8, Heckerman et al are silent regarding a step of storing additional digitized audio data for providing a further audio data stream synchronized with said video data stream.

Wakamoto teaches the storage of multiple streams of additional data synchronized with the video data stream (Col 3, lines 47-51 "The above storage medium has a first sound channel wherein in the practice area sound data is stored only in relation to the. specific sound, and a second sound channel wherein sound data is stored only in relation to sounds other than the specific sound").

As taught by Wakamoto, the storage of multiple channels of sound synchronized with a video data stream is well known, and provides for, among other features, separate tracks for each character.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify Heckerman et al in order to provide for multiple audio streams synchronized with the video stream.

7. Claim 9 is rejected under 35 U.S.C. 103(a) as being unpatentable over the combination as applied to claim 6 above, and further in view of Tanizawa et al (US PG Pub 2002/0030334).

Regarding claim 9, Heckerman et al are silent regarding gain control data adapted to control audio gain at selected times during a running of said synchronized audio and video data streams.

Tanizawa et al teach the storage of volume control data by an editing system (Paragraph 0420 “The fade-in gain coefficient string Kin (n) is multiplied to the fade-in data string Xin (n). This value increases linearly to 0 to 1”).

As taught by Tanizawa et al, gain control data is well known, providing the user with control over the loudness of various sound levels in the reproduction of audio data, and would therefore have been an obvious addition to Heckerman et al to one of ordinary skill in the art at the time of the invention.

8. Claims 19-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman et al and Nefian, and further in view of Wakamoto.

Regarding claim 19, Heckerman et al disclose an apparatus for processing audio and video data, comprising:

- means for selecting from data representing synchronously streamable video and audio data, scene data representing a portion of a stream of the streamable data and measuring durations of and intervals containing audible time-varying acoustic features of audio data within said data (Col 11, lines 39-2 "for a pointer to be inserted into the text and/or audio for synchronization purposes, the recognized text, bracketing the identified point of silence must have been correctly identified"); and
- means for populating a database with data and measurements provided by said selecting and measuring means (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Wakamoto teaches the storage of scene data (Col 12, lines 40-43 "The control data for use with playback control can be stored for example on track 1 immediately after the lead-in area, and comprises data for the purpose of displaying menu screen or jumping to a scene selected from that menu screen"), providing the user with a means of rapidly selecting the scene in which he or she desires to mimic the audio track.

As taught by Wakamoto, the storage of scene data is well known for menu purposes, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

Regarding claim 20, Heckerman et al disclose an apparatus comprising means for populating said database with text related to said scene data and measurements (Col 8, lines 14-16 "the speech recognizer module 312 generates a set of recognized text with time stamps from one or more audio files").

Regarding claim 21, Heckerman et al are silent regarding populating a database with still data representative of static video data extractable from said scene data

Wakamoto teaches the storage of still data for menus and role playing purposes (Col 15, lines 36-39 "Segment play data includes still data (c-f. FIG. 12) for use in menus for selecting role playing game or movie modes and sound or subtitle. Playback control data includes scene jumping data for use in role-playing games, together with subtitle and sound channel information for use at such times").

As taught by Wakamoto, the storage of still data is well known, and provides the user with a means of accessing various locations in the recording, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

9. Claims 22-24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman and Nefian, and further in view of Wakamoto.

Regarding claim 22, Heckerman et al disclose audio and video data processing software (Col 5, line 66 – Col 6, line 1 "The present invention will be described in the general context of computer-executable instructions, such as program modules, being executed by a personal computer") comprising:

- a database program adapted to populate a database with data and measurements provided by said selection and measuring program (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Wakamoto teaches the storage of scene data (Col 12, lines 40-43 "The control data for use with playback control can be stored for example on track 1 immediately after the lead-in area, and comprises data for the purpose of displaying menu screen or jumping to a scene selected from that menu screen"), providing the user with a means of rapidly selecting the scene in which he or she desires to mimic the audio track.

As taught by Wakamoto, the storage of scene data is well known for menu purposes, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

Regarding claim 23, Heckerman et al disclose audio and video data processing software according to claim 22, wherein said database program is further adapted to populate said database with text related to said scene data and measurements (Col 8, lines 14-16 "the speech recognizer module 312 generates a set of recognized text with time stamps from one or more audio files").

Regarding claim 24, Heckerman et al disclose audio and video data processing software wherein said database program is further adapted to populate said database with data representative of video data extractable from said scene data (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Heckerman et al are silent regarding populating a database with still data representative of static video data extractable from said scene data

Wakamoto teaches the storage of still data for menus and role playing purposes (Col 15, lines 36-39 "Segment play data includes still data (c-f. FIG. 12) for use in menus for selecting role playing game or movie modes and sound or subtitle. Playback control data includes scene jumping data for use in role-playing games, together with subtitle and sound channel information for use at such times").

As taught by Wakamoto, the storage of still data is well known, and provides the user with a means of accessing various locations in the recording, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

10. Claim 25 rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman et al and Nefian, and further in view of Schulze (4,918,730).

Regarding claim 25, Heckerman et al disclose a method of processing audio data comprising the steps of:

- deriving from first audio data first feature data representative of audible time-varying acoustic features of the first audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio

versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize");

- comparing said first and second feature data and determining timing differences between the first and second feature data (Col 11, lines 6-8 "matching recognized words or sequences of recognized words in the recognized text to those found in the text corpus");
- editing the second audio data in dependence upon said timing differences such as to provide edited second audio data having a synchronous relation to said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files"); and
- outputting synchronously said edited second audio data with video data having a synchronous relation to said first audio data (Col 11, lines 62-64 "File synchronization can be accomplished by, e.g., adding a bi-directional pointer linking the text and audio files, to the text files, and, optionally, to the audio files").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"),

but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Heckerman et al disclose a first audio data, but are silent regarding the treatment of a second audio data.

Schulze teaches the derivation of feature data of a second audio data (Col 1, lines 48-49 "The underlying principle of the invention is to compare the enveloope3s of signals that are being evaluated" and Col 2, lines 50-52 "a signal sequence which is to be examined can be compared with several stored signal sequences"), providing feature data representative of audible time-varying acoustic features of the second audio data with a low resolution signal.

As taught by Schulze, deriving feature data representative of audible time-varying acoustic features of a second audio data is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Heckerman is silent regarding the muting of the first audio data.

Wakamoto teaches the muting of audio playback for various original voices as the user's voice is played back (Col 7, lines 45-49 "Selecting channels in this manner causes only the sound signal [1] on CH1 to be played back as voice next time the tape is played back: the man's voice A is heard from the left-hand speaker 8 [L], while the woman's voice B is turned off").

As taught by Wakamoto, muting of original channels that have been replaced by new channels is well known, providing the user with the ability to isolate the new recording, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

11. Claims 26-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Heckerman et al and further in view of Wakamoto.

Regarding claim 26, Heckerman et al disclose a method of processing audio data, comprising the steps of:

- selecting scene data representing a portion of a stream of the streamable data (Col 11, lines 39-2 "for a pointer to be inserted into the text and/or audio for synchronization purposes, the recognized text, bracketing the identified point of silence must have been correctly identified");
- measuring durations of and intervals containing audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary

works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize"); and

- populating a database with data and measurements selected from and measured in the scene data (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Wakamoto teaches the storage of scene data (Col 12, lines 40-43 "The control data for use with playback control can be stored for example on track 1 immediately after the lead-in area, and comprises data for the purpose of displaying menu screen or jumping to a scene selected from that menu screen"), providing the user with a means of rapidly selecting the scene in which he or she desires to mimic the audio track.

As taught by Wakamoto, the storage of scene data is well known for menu purposes, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

Regarding claim 27, Heckerman et al disclose a method comprising:

- deriving from the audio data in the scene data feature data representative of audible time-varying acoustic features of the audio data (Col 11, lines 26-30 "locations in the recognized text where silence preceded and/or followed by correctly recognized words are identified. As discussed above, in the case of audio versions of literary works and other works read allowed [sic] and recorded for commercial distribution purposes, silence is often a particularly easy to recognize"); and
- populating the database with said feature data (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

Regarding claim 28, Heckerman et al disclose a method further comprising creating text data related to said scene data and measurements (Col 8, lines 14-16 "the

speech recognizer module 312 generates a set of recognized text with time stamps from one or more audio files") and populating said database with said text data (Col 9, lines 1-5 "The speech recognizer module 312, generates from the audio corpus 20 a set 406 of recognized text which includes time stamps indicating the location within the audio corpus of the audio segment which corresponds to a recognized word").

12. Claim 29 is rejected under 35 U.S.C. 103(a) as being unpatentable over the combination as applied to claim 26 above, and further in view of Nefian.

Regarding claim 29, Heckerman discloses the synchronization of audio to a text file as analyzed and discussed above, and notes that synchronization of audio and a display is done (Col 7, lines 54-57 "In the case of synchronized text and audio files, the computer system 120 can switch between audio and text presentation modes or simultaneously provide audio corresponding to text being displayed"), but does not specifically disclose the receipt of a video file synchronized with an audio file.

Nefian teaches the receipt of a video stream and a synchronized audio stream (Col 1, lines 54-56 "multistream HMM using assumed state synchronous audio and video sequences is used"), providing a base for manipulating the timing of the audio stream to meet the goals of the invention.

As taught by Nefian, the receipt of synchronized audio and video streams is well known, and would have been obvious to include in Heckerman et al to one of ordinary skill in the art at the time of the invention.

Heckerman et al are silent regarding populating a database with still data representative of static video data extractable from said scene data

Wakamoto teaches the storage of still data for menus and role playing purposes (Col 15, lines 36-39 "Segment play data includes still data (c-f. FIG. 12) for use in menus for selecting role playing game or movie modes and sound or subtitle. Playback control data includes scene jumping data for use in role-playing games, together with subtitle and sound channel information for use at such times").

As taught by Wakamoto, the storage of still data is well known, and provides the user with a means of accessing various locations in the recording, and would therefore have been an obvious addition to Heckerman by one of ordinary skill in the art at the time of the invention.

13. Claim 30 is rejected under 35 U.S.C. 103(a) as being unpatentable over Wark (7,243,062).

Regarding claim 30, Wark discloses graphical user interface software (Col 3, lines 61-62 "The method 200 is preferably implemented in the system 100 by a software program executed by the processor") comprising:

- a video and graphics display program adapted to control a display screen to display moving pictures in response to a stream of video data and to display a plurality of graphically defined control areas on said screen (Col 10, lines 66 – Col 11, line 7 "a media editor 800 within which the method 200 (FIG. 2) of segmenting a sequence of sampled audio into homogeneous segments may be practiced. In particular, the media editor 800 is a graphical user interface, formed on display 114 of system 100 (FIG. 1), of a media editor application, which is executed on the processor 105. The media editor 800 is operable by

a user who wishes to review recorded media clips, which may include audio data and/or audio data synchronised with a video sequence, and wishes to construct a home production from the recorded media clips");

- a control module adapted to detect selection of a said control area by coincidence of cursor positioning and actuation of a pointing device and to generate respective control signals in response to such selection (Col 11, lines 63-64 "The transition lines 822 illustrate borders of segments, such as segment 830"); and
- an output program adapted to respond to said control signals by outputting selected synchronized streams of video data and audio data (Col 11, lines 26-30 "The media clip(s) associated with the aforementioned selected icon(s) 804 are played from a selected position and in the desired sequence, in a contiguous fashion as a single media presentation, and continues until the end of the presentation at which point playback stops"),

Wark is silent regarding recording of the output of the editor.

The Examiner takes official notice that recording the output of video editors is notoriously well known, providing the user with a means to store and distribute his or her work, and would therefore have an obvious addition to Wark to those of ordinary skill in the art at the time of the invention.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to JAMES A. FLETCHER whose telephone number is

(571)272-7377. The examiner can normally be reached on 7:45-5:45 M-Th, first Fridays off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Miller can be reached on (571) 272-7353. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/John W. Miller/
Supervisory Patent Examiner, Art Unit 2623

JAF
14 August 2008